(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Karen J Thomson

Entitled
A Computational Study of the Mechanism for F1-ATPase Inhibition by the Epsilon Subunit

For the degree of Master of Science

Is approved by the final examining committee:

Jingzhi Pu
Chair

Haibo Ge

Rajesh Sardar

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Jingzhi Pu

Approved by: Eric Long 07/09/2013
Head of the Graduate Program Date

A COMPUTATIONAL STUDY OF THE MECHANISM FOR F1-ATPASE INHIBITION BY THE EPSILON SUBUNIT

A Thesis

Submitted to the Faculty

of

Purdue University

by

Karen J. Thomson

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

August 2013

Purdue University

Indianapolis, Indiana

For my parents.

ACKNOWLEDGMENTS

I would like to thank the members of the committee, Dr. Haibo Ge and Dr. Rajesh Sardar, and especially my advisor Dr. Jingzhi Pu, for their patience and support. I would like to thank the post-doctoral members of the group, Dr. Yan Zhou and Dr. Pedro Ojeda, for their support and advice. I would also like to thank the department administrative members of Kitty O'Doherty and Beverly Hewitt for all of their assistance during the program and especially leading up to the completion of this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

SYMBOLS

m	mass
r	radius
v	velocity
t	time

ABBREVIATIONS

ATP	Adenosine triphosphate
CA	Alpha-positioned Carbon
CTD	C-Terminal Domain
NTD	N-Terminal Domain
FACTS	Fast Analytical Continuum Treatment of Solvents
K	Kelvin, temperature unit
MD	Molecular Dynamics
MC	Main Chain (of nucleic acid)
PDB	Protein Data Bank, associated structural data file type
SC	Side Chain (of nucleic acid)
TMD	Targeted Molecular Dynamics
fs	femtosecond = $1.0x10^{-15}$ seconds
ps	picosecond = $1.0x10^{-12}$ seconds

NOMENCLATURE

1BSN	RCSB accession code for PDB structure of F_1-ATPase compact ε subunit from *E. coli*
3OAA	RCSB accession code for PDB structure of F_1-ATPase subunits $(\alpha\beta)_3\gamma\varepsilon$ with extended ε, from *E. coli*

ABSTRACT

Thomson, Karen J. M.S., Purdue University, August 2013. A Computational Study of the Mechanism for F1-ATPase Inhibition by the Epsilon Subunit. Major Professor: Jingzhi Pu.

The multi-protein complex of F_0F_1 ATP synthase has been of great interest in the fields of microbiology and biochemistry, due to the ubiquitous use of ATP as a biological energy source. Efforts to better understand this complex have been made through structural determination of segments based on NMR and crystallographic data. Some experiments have provided useful data, while others have brought up more questions, especially when structures and functions are compared between bacteria and species with chloroplasts or mitochondria.

The epsilon subunit is thought to play a significant role in the regulation of ATP synthesis and hydrolysis, yet the exact pathway is unknown due to the experimental difficulty in obtaining data along the transition pathway. Given starting and end point protein crystal structures, the transition pathway of the epsilon subunit was examined through computer simulation.The purpose of this investigation is to determine the likelihood of one such proposed mechanism for the involvement of the epsilon subunit in ATP regulation in bacterial species such as *E. coli*.

1. INTRODUCTION

All structures were visualized using VMD, unless originating from another source as noted. VMD: Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics" J. Molec. Graphics 1996, 14.1, 33-38 [1].

1.1 Structural properties

1.1.1 Properties of ATP synthase

The significance of Adenosine triphosphate (ATP) pertains to its universal use as an energy source in biological species, whether used by plants, animals, bacteria, or even archaea. As such, the protein responsible for its production is highly conserved but not identical across taxa. In order to advance the understanding of ATP regulation, numerous experiments have been reported to determine the structural properties and their corresponding functions. As a result, we now known that in common to all versions of ATP synthase are the key components that regulate the proton gradient across the cell membrane and channels it for use in the synthesis and hydrolysis of ATP as needed by the cell [2–9]. It is known that in the F_1 unit of ATP synthase, the epsilon (ε) subunit is stabilized in a compact form (following the convention of Cingolani et al. 2011, known as ε_C) when the concentration of ATP is adequate. However, at low concentration of ATP, it undergoes a change in conformation that enables it to inhibit the rotation of the central stalk (or γ subunit) and prevent hydrolysis of ATP. Since the direction of γ rotation is important, it is known from experiments [7, 8, 10] that the counter-clockwise rotation of γ is observed during hydrolysis, and the reverse direction for synthesis.

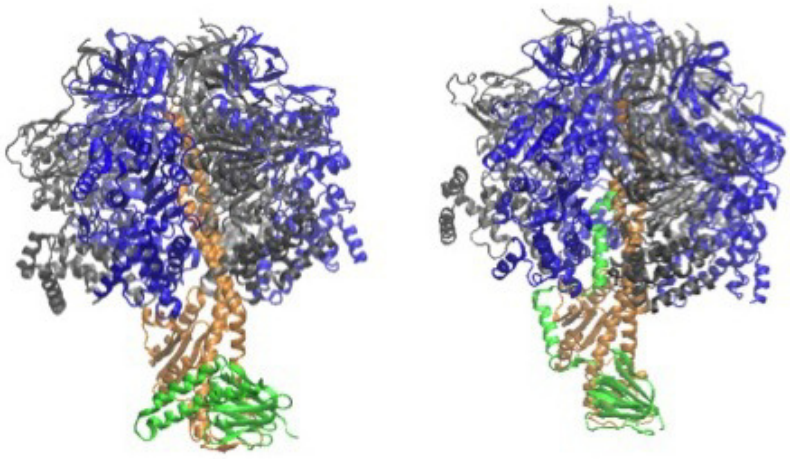


Figure 1.1. (Left) Structure of F_1-ATPase, crown subunits α (in gray) alternate with β subunits (in blue) while γ subunit is shown in orange. The ε subunit is seen in compact form, in green. (Right) Structure of F_1-ATPase, with ε seen in its extended form, in green

The work of Cingolani et al. 2011 [7] introduces a crystal structure for an extended conformation of the epsilon subunit (hereafter referred to as ε_χ) that inhibits rotation in the γ subunit. It remains to be investigated the likelihood of the proposed transformation, in addition to what exactly causes the change in conformation– in short, how does it work?

In general, the ATP synthase enzyme is a trans-membrane protein consisting of two main domains, designated F_0 and F_1. In bacteria, the F_0 domain acts as a proton channel, and regulates the concentration of protons and ions across the membrane. The F_1 domain is where the catalytic cycle of ATP takes place, which consists of synthesis and hydrolysis, depending on the proton concentration. [7, 8]

According to Muench, Trinick, and Harrison [9], three main ATP synthase complex types diverge from a distant common ancestor. While many residue sequence motifs are conserved among ATPases, they vary significantly in structure. The F_0F_1 - or F-ATPase of bacteria (i.e., *E. coli*) has the fewest number of components, with ab_2c-

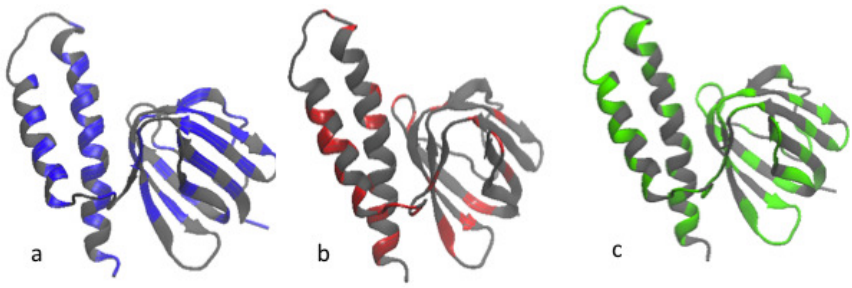


Figure 1.2. (a) Structure of ε_C, hydrophobic residues shown in blue against structure. (b) Structure of ε_C charged residues shown in red against structure. (c) Structure of ε_C, polar residues shown in green against structure. Structure was produced using the 1BSN crystal structure, using VMD [1, 3]

$\alpha_3\beta_3\gamma\delta\varepsilon$. The F_1-ATPase of bovine mitochondria is similar in structure to that of *E. coli* yet contains additional subcomponents with some unknown purposes.

In addition, the ε components of the bacterial and bovine mitochondrial ATPases do not overlap. The ε chain of *E. coli* corresponds to δ in bovine mitochondria (part of the rotor axle, participates in inhibition of ATP hydrolysis when H^+ chemical potential is low), while the ε component in bovine mitochondria (also found in the rotor axle) has no equivalent in *E. coli*. [7, 9]

According to Sekiya et al. 2009 and Futai et al. 2012 [8, 10], the DELSEED motif (residues 380-386 of the β subunit), which occupies the lowest outer loop, contains charged residues that are likely to interact with charged residues on the CTD helix of the ε subunit as part of a rotational inhibitory mechanism. Given the structures present in this study (see Figure 1.2), it is possible to investigate the interaction of the two subunits by the analysis of a simulation of the entire F_1 complex.

1.1.2 Properties of the epsilon subunit

There are several possible factors that may affect transitions between ε_C and ε_χ and are independent from ATP, Mg^{2+} and proton channel activity. These include the distribution of charged residues, hydrogen bonding and un-bonding (the binding

	B-strand (N-terminal)	Disordered region-loop 1	α- helix 1	Disordered region-turn 3	α-helix 2	Disordered region 3 – turn 5 (C-terminal)	α-helix 3	Disordered region 4
Residues: 1BSN	1:85	86:92	93:106	107:113	114:135	136 :138	N/A	N/A
Residues: 3OAA	1:82	83:84	85:102	103:111	112:126	127:128	129:133	134:138

Table 1.1
Mapping of secondary structures of minimized compact and extended segment ε, by residue number.

zipper model), and the distribution of hydrophobic/hydrophilic residues [11,12]. The most agreed upon mechanism is that of an Alanine (or Lysine) "zipper", where the hydrogen bonds between residues of parallel α-helices, as present in the compact form of the ε subunit, are formed or broken during transitions between the compact and extended forms of the subunit. It may be suggested that the concentration of the proton gradient may partially determine the rate at which long-range hydrogen bonds are maintained, in the absence of other causal factors. As part of this study, the hydrogen bonds of such residues will be monitored to determine the possibility of such an occurrence.

1.2 Progress on the epsilon subunit

This section contains a brief summary of the literature pertaining to the ε subunit in F_1 -ATPase, emphasizing important developments in structure and mechanism.

While most works focus on the γ subunit, a few select works focus on the role of epsilon in F-ATPase. From one of the first crystallographic studies in 1997, Uhlin, Cox, and Guss [2] made use of X-ray crystallographic data to resolve ε at 2.3 Å. The ε subunit was described as a β-sandwich consisting of the first 85 residues (the N-terminal domain), connected by 4 residues in a loose loop that connect to the remaining 46 residues forming "an antiparallel coil of two α-helices" (the C-terminal

domain, residues 90-136) that are linked by another short loop region, residues 107-112. The two domains appear to run anti-parallel to each other. It is thought that a hydrophobic region that connects the centers of the β-sheets is significant to interaction with γ. In the folded state of the α-helices, it was proposed that an alanine zipper forms between the outer residues of the first helix and the inner residues of the second helix, involving residues Ala 94 - Ala 117, Ala 97 - Leu 121, Ala 101 - Ala 124, and Ile 105 - Leu 128. It was also suggested that there is hydrogen bonding between specific residue side chains: Ser 10 O - Arg 93 NH1; Glu 70 OE2 - Gln 72 NE2; and Tyr 114 OH - Ser 118 O.

Also, it was suggested that Glu 70 interacts with residues Leu 42, Ile 68, and Leu 79 that also come into contact with the γ subunit during rotation. In a sequence alignment comparison with the bovine mitochondrial equivalent subunit, the alanine zipper is lost in the C-terminal domain, unless additional gaps are supposed, resulting in a looser structure in the bovine mitochondrial equivalent. Given prior studies in cross-linking cysteine substitutions, it was suggested that the ε subunit behaves independently of the rotation of the $(\alpha\beta)_3$ unit. Crystal structure atomic coordinates were submitted to the Brookhaven Protein Data Bank, accessible as 1AQT. [2, 13] Most importantly, the question arose as to what purpose the ε subunit serves.

In Wilkens and Capaldi 1998 [3], a solution phase NMR averaged structure was proposed for the ε subunit of F_1-ATPase in *E. coli*. The resolution for strong signals was given between 1.8-3.5 Å for a structure that appears similar to that offered by Uhlin, Cox, and Guss [2] but closer in packing between the N- and C-terminal domains. Specifically, there is close hydrophobic region interaction between the last length of the β-sandwich (labeled as strand β7) of the N-terminal domain and the C-terminal endpoint α-helix. The hydrophobic residues Ala 126, Gln 127, Val 130, Ile 131, Leu 133, Thr 134 and Ala 137 interact with NMR indicated counterparts in the N-terminal domain of Gly 48, Phe 61, Tyr 63, Gly 86 and Gln 87.

There is one reported caveat that possibly affected the structure, in that the ε subunit is known to operate at pH < 7.4, while the pH used to obtain the structure

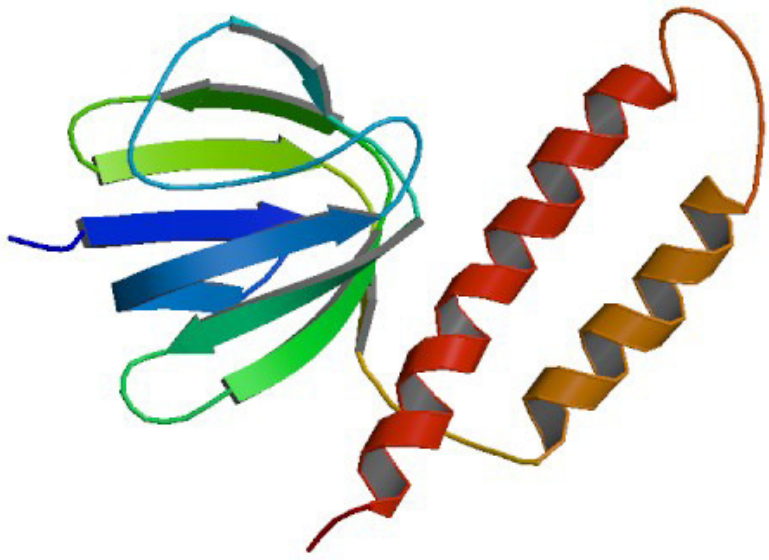


Figure 1.3. From Wilkens and Capaldi 1998, the structure of the epsilon subunit of F_1-ATPase from *E. coli* [3].

is well above that value in order to maintain a monomeric solution. At higher pH, the secondary structure of disordered regions is affected. Due to high pH and D_2O solvent exchange, many of the long-range signals between the side chains of the two α-helices appear to overlap. The measured NMR relaxation parameters from the main chain indicate conformational stability within the measured timeframe. The main advantages of the NMR study included the determination of the stability of the X-ray crystal structure, chemical bond information on interactions between nucleic acids, and additional interactions sites. The reported average NMR structure atomic coordinates were submitted to the Brookhaven Protein Data Bank, accessible as 1BSH. [3, 13]

The second concern of Wilkens and Capaldi [3] was of resolving the number of conformations involved in the process of ATP/ADP catalysis in F_1. Given the most likely states of ε, the number of rotational steps involved in the catalytic cycle by $(\alpha\beta)_3$ could be determined. An analysis of the fragments produced by cleaving with trypsin indicated a minimum of two states, an ATP (AMP-PNP+Mg^{2+}) form and an ADP (ADP/Mg/P_i) form bound to specific sites in two out of three β subunits simultaneously. Since trypsin digestion is likely to produce more fragments when the C-terminal of the ε subunit is linked to α and β subunits than when the C-terminal α-helices are blocked by the N-terminal β-sandwich, the outcome would determine that more fragmentation would occur with ε bound to α or β subunits in the ATP form but not for the ADP form. The digested form of ε alone showed as few fragments from ε as the digested form of the $\varepsilon - F_1$ complex with ADP present, and a distinct lack of fragments observed in the ATP form. Thus, it appears that the reported X-ray and NMR atomic structures belong to the ADP form. Additionally, previous cross-linking experiments show binding of ε along the C-terminal hairpin to α and β subunits in ATP form.

Further investigation using a mutant version of the ECF_1 complex with two-step sequential Cys 138 crosslinking yielded a product with a mass that corresponded to a $\beta - \varepsilon - \beta$ complex. Moreover, this allowed for specific designation of each β according

to its position: β_{TP} to indicate binding to AMP-PNP, β_{DP} to indicate binding to ADP, with the third β (β_E) available. However, it was suggested that not only does the $\gamma-$ complex rotate in three steps with each β subunit in the transition from ATP to ADP, but that $\gamma-\varepsilon$ also switches between each α subunit (ADP form) to the ATP state. This would extend the rotations to six, one for each subunit of the c-ring in F_0 if ε were required to adjust and relink with each rotation. Lastly, given ADP-ATP dynamics, it is suggested that the form of ε studied is the more energetically favored configuration. [3]

In 2000, Rodgers and Wilce [4] reported a crystal structure for the $\gamma-\varepsilon$ complex at 2.1 Å resolution. The report related the influence of the ε subunit on the coupled rotation of γ between the F_0 and F_1 units. In addition to providing more detail about what is known about the F_0 unit, an intermediate position for ε is proposed. This position assumes an orientation that is an approximately orthogonal to that shown in the previously submitted structures. From the position of a downward pointed α-helical hairpin, the C-terminal helix is depicted at a right angle to the first helix, so that the disordered loop region between them extends outwards with an elbow-like appearance. It is this position that allows interaction with the α, β, and γ (denoted as γ' by Rodgers and Wilce [4]) subunits.

It was reported [4] that the β-sandwich of the N-terminal domain does not change position during the transition of the α-helix hairpin from folded to hinged state. While the entire amino acid sequence of ε is known, the structure presented does not show the residues that comprise the disordered loop regions, mainly 104-111. An alignment of coordinates for the crystallographic structures of previously reported mitochondrial bovine $F_1(MF_1)$ and the determined structure of γ- ε was performed using the brute force option of LSQMAN, and compared with the results using CLUSTAL X (graphical version of CLUSTAL W). The reported overlap between Cα structures was 34%, which suggests that there is considerable disagreement. However, based on the alignment results, it was deduced that there is little binding affinity between ε and $(\alpha\beta)_3$, while there is far more influence on c_1 by ε.

Additionally, more weight was given to previously reported cross-linking data to suggest that rotation depends more upon the γ subunit with little or no involvement from the ε subunit. Overall, the view of the role of the ε subunit had changed dramatically from direct inhibition of catalysis in F_1 to regulating $[H^+]$ and the rate of c subunit rotation due to its linkage between F_1 and F_0 . [2–4]

The reported crystallographic structure atomic coordinates were submitted to the Brookhaven Protein Data Bank, accessible as 1FS0 [4, 13].

The following year (2001), Tsunoda et al. reported the results of additional Cys residue cross-linking studies that used the ε and γ subunits in *E. coli* as a model for the δ and γ subunits in bovine mitochondria. The work investigated the purpose and role of the two known configurations for ε, in relation to the F_0 c-subunits and $(\alpha\beta)_3$ in F_1. The closed-form of ε replaced Ala 117 and Gln 42 of one c subunit, while the open-form of ε replaced Ser 118 and Lys 99 of the gamma subunit. The high yields of cross-linked products indicated the ability of ε to switch between the two states. Further investigation of this switch included the use of reductant DTT (dithiothreitol; (2S,3S)-1,4-bis(sulfanyl)butane-2,3-diol) to reverse the effects of cross-linking, and the F_1-ATPase inhibitor DCCD (N,N-dicyclohexylcarbodiimide). The use of DTT reversed the effects of cross-linkage and restored hydrolysis function, while DCCD disrupted both synthesis and hydrolysis functions, with and without cross-linkage. While the cross-linked open form of ε was disrupted by DCCD, the hydrolysis function measured significantly less activity than that of the synthesis mode. [5]

With studies of the epsilon subunit interacting with the DELSEED motif found in the β subunit, it is demonstrated that the C-terminal of the ε open-form plays a major role in the regulation of F_1 activity. Further, when the C-terminal was cross-linked into a fixed closed form, hydrolysis was hindered while synthesis continued. It was proposed that the direction of $(\alpha\beta)_3$ rotation while interacting with the ε subunit is favored during ATP synthesis, while the opposite direction relating to hydrolysis is sterically hindered at the release of the ε subunit. Tsunoda et al. reinstate rota-

tional steps at 120° for a three step process while not specifying how the transition relates to the timing of ε in a closed or open state in relation to the position of α or β subunits. Tsunoda and colleagues proposed that the role of the IF_1 protein in mitochondrial ATPase is limited to regulation of hydrolysis as controlled by pH or the proton gradient, leaving the role of δ (or ε) to regulate ATP synthesis through its ratcheting effect. [5]

The work of Yagi et al. 2007 [6] approaches the problem from a different direction. Instead of using *E. coli* or animal F_1F_0, ε from thermophilic *Bacillus subtilis* was used. Additionally, the problem was defined as how ATP affects the configuration of ε, as opposed to how ε affects the hydrolysis or synthesis of ATP. The determined structure of the ε subunit from the thermophilic bacteria was described almost identically to that of *E. coli*, with the exception of the angle of orientation for the N-terminal domain (NTD) and less flexibility at the C-terminal domain (CTD) hinge.

Based on crystallographic data of *E. coli*, a location for ATP binding to the closed form of the ε subunit was determined between residues Ile 88, Asp 89 and Arg 92, Ala 93 of the α-helix hairpin in the CTD, via hydrogen bonding. This forms the motif I(L)DXXRA (for X is any amino acid). Additional Arg residues form possible ATP binding pockets.

A study of the solution structure surmised that while ATP is not bound, the CTD is flexible compared to the NTD, especially in the loop regions. The ATP bound form maintains a rigid CTD as well as NTD. In order to reconcile the cross-linking studies with γ, it was proposed that with the expansion of the lower region of γ, the CTD of ε fully extends to align with the top portion of γ. The presence of ε prevents the rotation of γ during ATP hydrolysis due to steric hindrance, while rotation in the opposite direction (for ATP synthesis) is unhindered. The similarity of the *Bacillus sp.* structure to that of *E. coli* provides a reasonable basis for transference of a mechanism and structural properties. Given this transference, one compilation of multiple NMR structures for the α-hairpin indicates an unusual degree of flexibility at the disordered loop region separating the two helices. It is supposed that this

allows for a 90° rotation between the two helices that places one helix in front of the other, as opposed to a scissor-like expansion in the same plane.

The atomic coordinates for two NMR solution structures and one crystallographic structure were submitted to the Brookhaven Protein Data Bank, accessible as 2E5T, 2E5U, and 2E5Y. [6, 13]

The most recent structure was offered by Cingolani and Duncan (2011). The new crystallographic structure offers the most complete view of EF_1 so far, with $\alpha_3\beta_3\gamma\varepsilon$ subunits, and Mg, ADP, SO_4^{2-} ligands. The structure was deposited with the Protein Data Bank, accessible as 3OAA. Here, previous ambiguities are clarified: (1) the origin of proton motive force on the c-ring determines the direction of catalysis between synthesis and hydrolysis (2) the addition of a subunit outside $(\alpha\beta)_3$ connects F_1 with F_0, however it has yet to be discerned structurally (3) more precise roles of α and β subunits in the presence of specific ε intermediates during catalysis (4) support for the theory of the ε subunit as a molecular ratchet (5) interactions between γ, β_1 and the extended state of ε that produce differing rotational steps across types of ATPase. [7]

The contribution of most interest is the additional structural details of ε with γ. Here, ε c is the term favored to define the compact (or uninhibited) state, while the extended state is denoted as ε_χ. While there is little change in the compact form, the extended state is shown to differ from the compact in that parts of the α-helices become disordered loop regions, along with the formation of a "hook" at the end of the CTD.

Based on the trypsin digestion experiments and the additional structural detail on ε, Cingolani and Duncan offered a more in depth explanation to reconcile the evidence.

1.3 Importance of the epsilon subunit

The multi-protein complex of F_0F_1 ATP synthase has been of great interest in the fields of microbiology and biochemistry, due to the ubiquitous use of ATP as a biological energy source. Efforts to better understand this complex have been made through structural determination of segments based on NMR and crystallographic data (i.e., Uhlin et al. 1997). [2] Some experiments have provided useful data, while others have brought up more questions, especially when structures and functions are compared between bacteria and species with chloroplasts or mitochondria. [9]

Computational simulations have been used to realize several key mechanistic steps in the F_1 subunit [11]. Previously reported efforts to propose a partial mechanism for the role of the epsilon subunit were based on reported endpoint NMR and crystal structures, protein digestion with enzymes, and cross-linking studies (e.g., Uhlin et al. 1997; Wilkens and Capaldi 1998; Tsunoda et al. 2001; Yagi et al. 2007; Cingolani and Duncan 2011; etc). [2, 3, 5–7] Few of these structures are in agreement, nor do they indicate more than one intermediate stage due to the difficulty of obtaining intermediate structures of pure specimens experimentally. Additionally, the proposed mechanisms remain in competition with each other without further investigation of the interaction between the epsilon subunit and subunits in F_1 and F_0 at intermediate steps. Thus, a computational study of the epsilon subunit would be beneficial in the theoretical determination of its role in the ATP catalysis cycle.

One goal of this investigation is to determine an energetically favorable mechanism for the involvement of the epsilon subunit in ATP regulation in bacterial species such as *E. coli*. The first study of the ε subunit determines the intrinsic properties of the subunit itself, isolated from the rest of the complex. The second study identifies and quantifies properties that rely on the gamma subunit without rotation. The third and last study will involve all subunits, $\alpha_3\beta_3\gamma\varepsilon$. All three scenarios are performed without ADP/ATP, while rotation is neglected in the first two studies. This is in contrast to computational studies performed with rotation but without epsilon, that focus on the gamma component (e.g., Ma et al. 2002). [11]

2. THEORY AND COMPUTATIONAL METHODS

Overall, the goal of molecular mechanics is to simulate a collection of atoms in order to predict the position and trajectory of the collection over time. This provides the advantage of being able to determine physical characteristics of a chemical system without the limitations of real life experiments, such as data collection rates and imaging technology. In a simulation, if the starting masses and positions are known for a collection of atoms, then a program may calculate the new positions for the atoms at a new time, t'. The forces which comprise the complete potential energy of the system determine the resulting trajectory calculated at each time step. [14–17]

The next section relates the derivational basis for the numerical methods used in CHARMM, as CHARMm is the software primarily used in this study, and is comparable to most other methods it has preceded. Since the problem involves an ensemble of a large number of atoms, it is reasonable to treat the problem as classical, and ignore primary quantum effects. Secondary quantum effects that influence atoms in proximity to each other may be approximated as boundary condition potentials (i.e., van der Waals, Lennard-Jones, hydrogen bonding, non-bonding effects, etc) and are added to the overall potential energy of the system. [14, 16–19]

2.1 Molecular Mechanics

Classical theoretical mechanics provides the basis for computational simulations of molecular dynamics. From Newtons laws regarding the behavior of particles in motion, one may derive the working principles. Given the classic equation $F = ma$, one may define the variables such that

$$\vec{\mathrm{F}} = -\nabla\Phi \tag{2.1.1}$$

where the vectorized force is equivalent to the gradient of the potential. The full potential (as used by CHARMm [16]) is comprised of a sum of as many of the most significant forces that act upon each particle:

$$\begin{aligned}
&\Phi = \textstyle\sum_{bonds} K_b(b - b_0)^2 \\
&+ \textstyle\sum_{angles} K_\theta(\theta - \theta_0)^2 \\
&+ \textstyle\sum_{Urey-Bradley} K_{UB}(S - S_0)^2 \\
&+ \textstyle\sum_{dihedrals} K_\phi(1 + cos(n\phi - \delta)) \\
&+ \textstyle\sum_{impropers} K_\omega(\omega - \omega_0)^2 \\
&+ \textstyle\sum_{\substack{non-bonded \\ pairs}} \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \\
&+ \textstyle\sum_{residues} U_{CMAP}(\phi, \psi)
\end{aligned} \tag{2.1.2}$$

where each term contains a spring constant K pertaining to each force, and the distance from the equilibrium value, with the exception of the non-bonded term. The second to last term includes the terms for Coloumb and Lennard-Jones factors. Each sum is given in terms of spherical or rotational coordinates, e.g. r may be substituted

for b in the first term for bonds, and terms ϕ, ψ, ω are used for rotation. In the Urey-Bradley term, S is substituted in place of angular coordinates to represent the $1,3$ non-bonding dihedral connectivity.

Since the gradient of the potential acting on a system of particles may be described as:

$$\sum_{\substack{j=1,\\ j\neq i}}^{N} F_{ij} = m_i \, \frac{d^2 \, \vec{r_i}}{dt^2} \tag{2.1.3}$$

so that the sum of all force components acting on a system of particles determines the trajectory of the system of particles. Computationally, the computer program must integrate the differential terms to determine the position of each particle with respect to time. A Taylor series expansion is used to approximate derivatives in Euler form (to solve for $i \rightarrow i+1$), and integration becomes the process of solving multiple ODEs. Using the Explicit Central Difference Method as an example,

$$\frac{d^2 r_i}{dt^2} = \frac{r_i(t+\Delta t) + r_i(t-\Delta t) - 2r_i(t)}{2\Delta t^2} \tag{2.1.4}$$

the calculation is considered fourth order ($O(\Delta t^4)$). To simplify and perform the calculation, the differential form is integrated so that the explicit calculation is reduced to second order ($O(\Delta t^2)$) at the calculation of velocity. The length of the time step is chosen by the user, and is required to be smaller than the actual physical process. Computationally, reducing the order of the equation makes the calculation less expensive. However, the calculation becomes more expensive with decreasing time step size. [18]

$$r_i(t+\Delta t) = 2r_i(t) - r_i(t-\Delta t) + 2 \, \Delta t^2 \left(\frac{F_i}{m_i}\right) \tag{2.1.5a}$$

$$v_i(t) = \frac{dr_i}{dt} = \frac{r_i(t+\Delta t) - r_i(t)}{\Delta t} \tag{2.1.5b}$$

Since the Explicit Central Difference Method is also known as the Verlet Method (after Verlet 1967; see also Hockney 1970, Cuendet and van Gunsteren 2007) [20–22], it is sometimes described by the number of time steps (for Δt at $i,\ i+1,\ i+2,\ \cdots$) by the following

$$r_{i+1} = r_i + v_i\ \Delta t + \frac{1}{2}a_i \Delta t^2 + O(\Delta t^3) \tag{2.1.6a}$$

$$v_{i+1} = v_i + \frac{1}{2}(a_i + a_{i+1})\Delta t + O(\Delta t^3) \tag{2.1.6b}$$

which is fine, with the exception that if it is fully reversible, it cannot be self-starting and requires a different starting algorithm. If the time step progression is altered to half-steps (i.e., $i+\frac{1}{2},\ i+1,\ i+\frac{3}{2},\ \cdots$), the Verlet equations become

$$r_{i+\frac{1}{2}} = r_i + \frac{v_i}{2}\ \Delta t \tag{2.1.7a}$$

$$v_{i+1} = v_i + a_{i+\frac{1}{2}}\ \Delta t \tag{2.1.7b}$$

$$r_{i+1} = r_{i+\frac{1}{2}} + \frac{1}{2}v_{i+1}\ \Delta t \tag{2.1.7c}$$

and subsequently:

$$r(t+\Delta t) = r(t) + v\Delta t\ (t + \frac{1}{2}\Delta t) \tag{2.1.8a}$$

$$v(t+\frac{1}{2}\Delta t) = v\ (t - \frac{1}{2}\Delta t) + a(t)\ \Delta t \tag{2.1.8b}$$

which is known as the Midpoint Method, and the basis of the Leap Frog integrator. It has the advantages of having higher order errors cancel out over time, and is fully reversible in solving N-body problems. Given the advantages of the Midpoint Method over the original Verlet, the default dynamics integration method is now Leap Frog. However, the Verlet method may still be used if specified in the dynamics input of CHARMm. [16]

2.2 Molecular Dynamics

In Molecular Dynamics (MD), a simulation may be considered an experiment conducted in silico. First the coordinates for the structure of interest are constructed or obtained from a repository, such as the RCSB Protein Data Bank in studies of biomolecules. The raw file must be refined such that information is compartmentalized. For example, the preparation of a raw PDB file will contain a list of residues, the atoms present in each residue, structural connectivity, accompanying ligands, etc. The use of a smaller PDB file with only the list of atoms and their coordinates, combined with a PSF file that gives topology information, a CRD coordinate file for reference positions is applied with general files of topology and parameters that act as a lookup table of data to create the force field associated for the whole structure. After the consideration of solvent (for a protein system), there are stages for system energy minimization and equilibration. [16]

The use of force field dynamics has the advantages over other methods such as *ab initio*/Monte Carlo or semi-empirical methods in that it is more computationally efficient, can handle systems in solvent, more complex/higher atom number systems, as well as a greater range of organic, inorganic, polymer and protein systems. However, it does not consider electrons, or excited states [17], unless the use of more recent features are employed towards analysis of potential spectra. [16]

The second step in preparing a structure for study in an all-atom simulation is that of adding solvent. One method is to explicitly add water molecules for a determined volume, that may interact with the protein in terms of hydrophilic and hydrophobic interactions. However, since it is significantly more computationally expensive and time consuming to run a simulation that includes possibly more water molecules than protein molecules, methods have been developed wherever possible to produce the effect of solvent being present without having to consider the actual molecules. Therefore, the solvent effect is implied (or, implicit) by the representation of another force field that is added to the overall potential. [16, 23, 24]

An array of available implicit solvent methods are available, mostly based on Poisson-Boltzmann Continuum Electrostatics, or PBEQ. The PBEQ is considered the most accurate yet resource expensive method. It considers the free energy as the difference between the charge based reaction field in vacuum vs in a dielectric solvent, and imposes Dirichlet fixed potential boundary conditions. [16] The Generalized Born Electrostatics, as used with the GBMV and GBSW methods (Molecular Volume, SWitching respectively) that is similar to that of PBEQ, but uses pairwise interactions to integrate over the participating charges using the calculated effective Born radii. The method used in this study, the Fast Analytical Continuum Treatment of Solvation, or FACTS, is based on the GBSW method, but relies on the spatial symmetry of the displaced solvent using an empirical volume formulae. [16, 24]

2.3 Targeted Molecular Dynamics

While MD simulations are used in the process of equilibration, the resulting trajectory is that of a random walk restrained by interatomic connectivity. If a simulation was left to run indefinitely, eventually a transition may take place. To improve efficiency, methods have been developed to speed up the process of modeling transitions and reduce random activity.

Towards the investigation of molecular transition pathways, there are several methods available using CHARMm. According to Huang et al. 2009, [19] there are three main methods that differ by application of restraint. The method of BMD (Biased Molecular Dynamics) applies an adiabatic bias potential to enforce the movement of the starting structure towards the target, but with minimal restraint. The calculation is based on the increase of the progress variable, $\rho_0(t)$ and the applied potential, $H(\rho)$. In a sense, the potential acts as a leash to prevent completely free motion. However, this method is not suitable for determining the most energetically favorable pathway.

The second method is SMD, or Steered Molecular Dynamics. This method is usually compared with the technique of Atomic Force Microscopy (AFM), and used to simulate AFM experiments. The SMD method uses a harmonic force restraint, so that the pathway progression is independent of the target molecular structure, and entirely driven by the applied force towards the endpoint. In contrast to BMD, the applied restraint of SMD provides the most influence on the resulting pathway. Since the goal of this study was to determine a stable and energetically favorable transition pathway, neither of these methods were used. Instead, the method of Targeted Molecular Dynamics, or TMD was employed to study the transition between two given conformations of the epsilon subunit. [19]

In contrast to these two methods, is the method of Targeted Molecular Dynamics (or TMD) to determine the most energetically favorable trajectory between the two subunit configurations of interest. In the TMD method, a holonomic constraint is used (as opposed to an applied force restraint) to calculate the system Hamiltonian (in other systems, a Lagrangian) in terms of the time dependent parameter $\rho_0(t)$, the root mean square distance (RMSD).

$$H(\rho) = \begin{cases} (\frac{\alpha}{2})(\rho - \rho_0)^2 & (\rho < \rho_0) \\ 0 & (\rho \geq \rho_0) \end{cases} \tag{2.3.1a}$$

$$\rho_0(t) = \begin{cases} \rho_0(t - \Delta t) & (\rho < \rho_0) \\ \rho_0(t) & (\rho \geq \rho_0) \end{cases} \tag{2.3.1b}$$

Every time step progresses along a set fraction of the total RMSD between the starting and target structures, so that the size of the allowed RMSD step (MSID, or Mean Step Internal Deviation) determines the total number of time steps required to reach the target structure. Outside of this constraint, the trajectory follows the path of least energetic resistance. In effect, small corrective perturbations are made along the trajectory, as opposed to large periodic adjustments after a large deviation. This

advantage allows for the use of energy sampling methods, such as umbrella sampling, to construct progress maps of energy vs transition pathway progress. [16, 19, 23]

3. TMD OF EPSILON SUBUNIT

The study began with the investigation of the isolated epsilon subunit. Without the presence of other ligands or protein subunits, the results would indicate which properties were intrinsic to the epsilon subunit itself and do not require interaction with other subunits. While the study may be used as a control, it also simplified the analysis of the epsilon subunit. The results of the TMD simulations were then used to direct the advancement of the investigation.

3.1 Experimental

The all-atom simulation of the epsilon subunit was set up in three main stages consisting of (1) setup of initial PDB, PSF, and CRD (structural data) input files using FACTS22 [24] implicit solvation (2) molecular dynamics heating from 100 K to 300 K, followed by equilibration, and (3) TMD [25] trajectories of the transition from ε_C to ε_χ, and the reverse. The reverse transition of ε_χ to ε_C was subject to the same conditions and analyses as the forward transition, ε_C to ε_χ.

The initial setup began with downloading the formal crystallographic and NMR structure data from the Protein Data Bank website [13]. The structural data for 1BSN was collected by NMR, which provided data for hydrogen present in each residue. However, only data for the epsilon subunit was available. For study of this epsilon configuration with other F_1 subunits, those subunits were borrowed from the 3OAA and aligned to fit in place of the original epsilon configuration. The structural data for 3OAA was collected by X-ray crystallography, and contains subunits $\alpha_3\beta_3$ as a hexamer. Only chain H (corresponding to the first epsilon subunit) was selected.

The program CHARMm [15,16] versions c36a1 and c35b3 were used in each stage of input preparation and molecular dynamics equilibration, simulation, and most data analyses. The preparation of initial structural input files began with the selection of all atomic data for 1BSN chain A and 3OAA chain H (corresponding to the first epsilon subunit) as separate PDB files. Occurrences of the residue HIS were converted to HSD as opposed to HSE to reflect the protonation state at a physiological pH between 7.0 - 7.5. This range also matches the conditions under which solution and crystallographic data were collected, thus it was maintained throughout the simulation. Additionally, initial preparation included adding hydrogen to residues in crystallographic data, patching N- and C-terminals, generation of a PSF and CRD initial input files. Further preparation included the setup of implicit solvation using the method of FACTS22. A preliminary investigation of the system used the implicit solvation method of GBSW. [26] This method was found to be fairly efficient during initial setup, but comparably less so in molecular dynamics equilibration.

Once the initial input files for 1BSN and 3OAA were generated, a series of molecular dynamics steps were used to heat up the molecule to 300 K and then equilibrate to a sufficiently minimized configuration. The heating step consisted of three stages, 0 - 100 K, 100 - 200 K, and 200 - 300 K. Each stage consisted of loading CHARMm topology and parameters files, reading the initial structure files, setting up FACTS22 implicit solvation, creating output files, defining harmonic constraints on the protein itself, and the SHAKE [27] constraint on fixed hydrogen bond lengths– all before the first time step in simulation. These parameters and those for the dynamics portion were kept almost completely constant throughout the equilibration sequence. An exception to this was the decrease of the harmonic constraint factor from 10.0 to 1.0 over five stages.

Each dynamics run was set to run at an interval of $1.0\ x\ 10^4$ time steps, with each time step equivalent to 2 fs. The number of time steps was increased to $3.0\ x\ 10^4$ after three stages of heating and five stages of equilibration, for an additional five stages and a total time of 460 ps prior to TMD trajectories. Each step following the

first heating continued the results of the previous stage. All dynamics simulations after heating were conducted at a constant temperature of 300 K using the canonical ensemble. These steps were each performed on 1BSN and 3OAA separately. The results of minimization were combined for use in TMD. The trajectory output files were inspected to verify the final configuration prior to use in TMD.

After minimization was reached, the restart information from the final minimization run was used to begin TMD. The input for TMD was nearly identical to that used in the early equilibration steps, with minor exceptions. Those exceptions included the selection definitions of both 1BSN and 3OAA initial reference structures, specifying both chains as part of the protein, introduction of the nucleic acid backbone atoms, adjustment of SHAKE to fast (i.e. the default settings limiting the iterations to 500 and the tolerance set to $1.0\ x\ 10^{10}$), and the addition of TMD specific parameters. Once parameters were defined, each was maintained for each step. A total of four stages were performed with a total trajectory time of 160 ps to reach RMSD between the starting and target structures of 1.5 Å. The total computational runtime (not counting analysis) required $\tilde{1}1.6$ hours of parallel processing using 8 nodes on the Big Red cluster at Indiana University.

3.2 Results and Discussion

3.2.1 Analysis of TMD

An analysis of the TMD trajectories included the determination of the dihedral angles of the loop segment joining the two helices, helical crossing angles, select atomic distances, and the frequency of hydrogen bonding between the two helices, each with respect to time. An examination of the calculated dynamics energies (see Table 3.1) for the isolated epsilon subunit compact form and the extended form indicates that an energetically minimized compact form is the more favored configuration. Additionally, the reverse TMD path from the extended to compact form is nearly identical, albeit less stable, indicating that additional energy or structures are required to make the

DYNAMC> Averages for the last 5000 steps:

AVER DYN: Step	Time	TOTEner	TOTKe	ENERgy	TEMPerature
AVER> 5000	620.00000	-514.70323	1565.81513	-2080.51835	300.00153
AVER> 5000	620.00000	-437.45986	1566.09476	-2003.55462	300.05510

Table 3.1
Average energies for final stage of 1BSN, 3OAA. First entry shows energetics for 1BSN to 3OAA, second entry shows 3OAA to 1BSN, TMD after 20 000 timesteps.

transition fully reversible. (i.e., fraying of helices in the extended form TMD into lengthened disordered regions) Here, the role of other agents affecting structure and energetics is indicated, even if the function of such deviations is not evident. This is consistent with proposed mechanism for regulation of ATP use; the presence of ATP or other channel activity is required to remove the molecular inhibitor, and when absent, returns to the inhibited configuration.

In TMD simulations, the transition from the 1BSN (compact) configuration to that of 3OAA (extended) follows several steps of gradual separation between the β-sheet and α-helix segments, and extension of the α-helices. Initially, the α helices are parallel to each other, while the C-terminal helix is roughly parallel to the β-pleated N-terminal domain (Figure 3.1 (a)). This allows for extensive hydrogen bonding between residues of both domains in close proximity. When these bonds break, the two domains are weakly separated by a disordered loop region. Further separation occurs when the C-terminal end of the second α-helix swings outward, twisting the first α-helix along with it (Figure 3.1 (b, c)). As the C-terminal end pulls away, the angle between the second loop region joining the two helices and the N-terminal domain increases (Figure 3.1 (d)). While the distance and angle between the first helix and the N-terminal domain increase, the two helices gradually twist apart (Figure 3.1 (e-g)). When the angle between the two helices is nearly 180°, the disordered loop

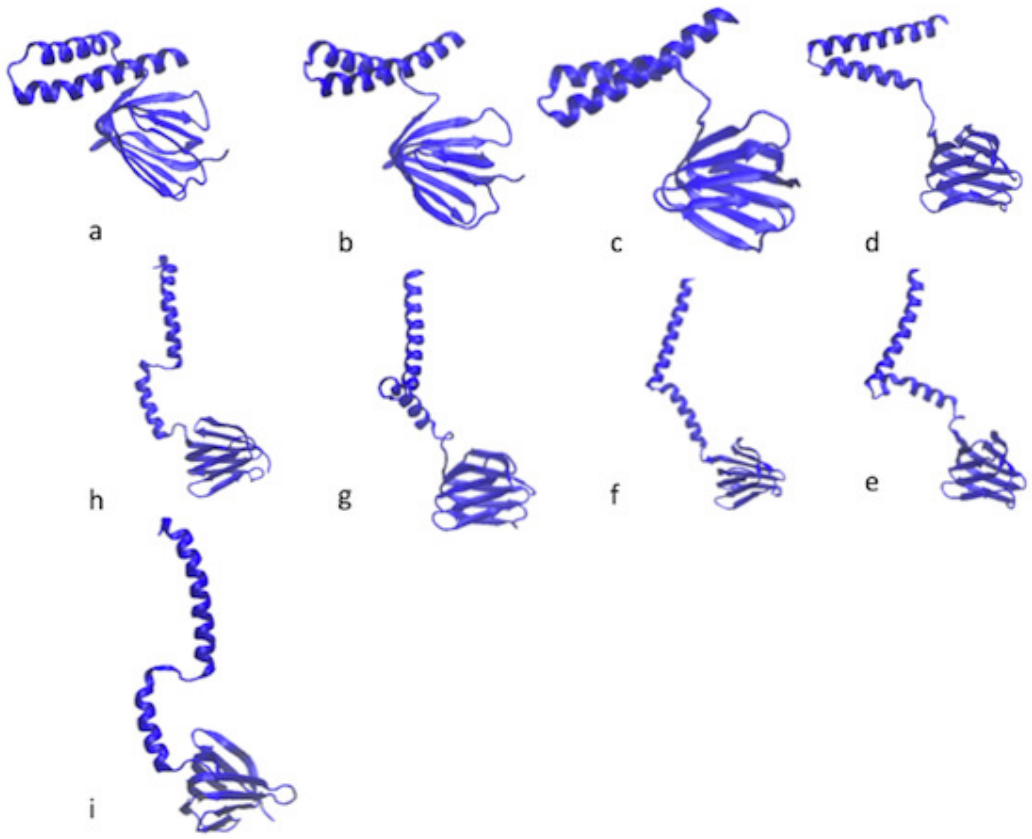


Figure 3.1. TMD simulation of transition from 1BSN to 3OAA conformations. Initial structure at (a) 2 ps (b) 25 ps (c) 50 ps (d) 80 ps (e) 95 ps (f) 110 ps (g) 120 ps (h) 150 ps (i) 160 ps; final structure

regions extend to allow the helices to rotate into more balanced positions until the configuration of 3OAA is reached (Figure 3.1 (h-i)).

These groups occur between residues at three distinct regions. The first region to disengage are the ends of the helices furthest from the hinge, e.g. residue 91 from residue 135 (see Figure 3.2). The second region to separate is e.g. residue 102 from residue 120. Lastly, the residues to each side of the hinge separate, e.g. 104 from 114.

The calculation of the dihedral angles for the Loop 2 (or hinge) region between the two helices was performed to provide information about the overall protein characteristics, allowed residue backbone torsion angles, and overall flexibility of the hinge. This enables a better understanding about the subunit movement as a whole.

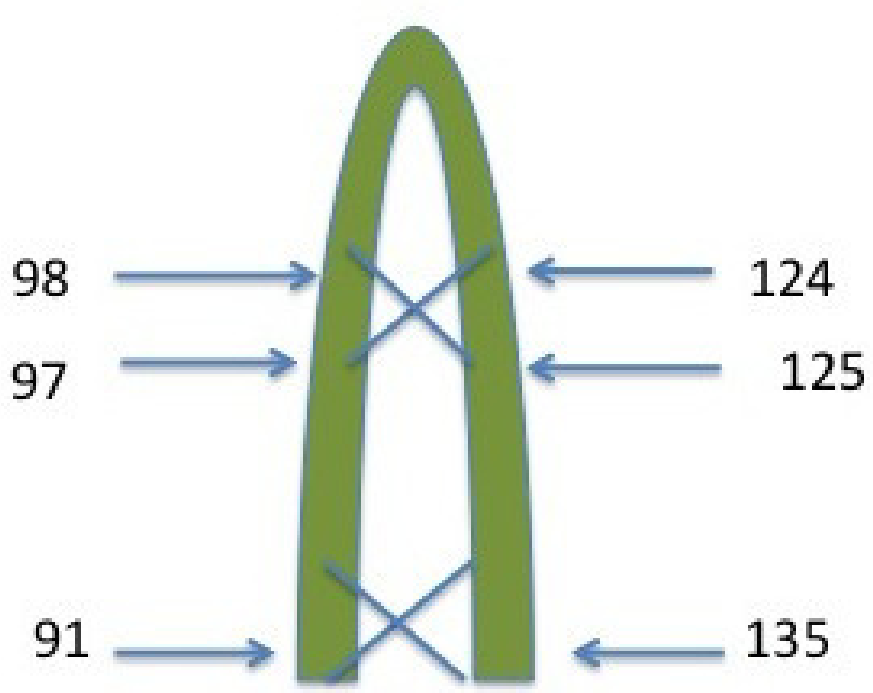


Figure 3.2. Cartoon schematic indicating relative positions of selected atomic pairs for distance calculations over transition trajectories

Since the expected conformational change requires a near 180° change with respect to the relative positions of the two helices, it is expected that the range required for this change to take place is not prohibited by the torsional ranges of the hinge residues. The structural data file includes the identities of hinge residues 107-108-109-110-111 as Serine (SER)-Serine (SER)-Histidine (HSD)- Glycine (GLY)- Asparagine (ASP). A rubric for disallowed regions (due to steric hindrance) was given by Pal and Chakrabarti 2002 [28], while Hovmöller et al. 2002 [29] surveyed residue types with allowed regions and their interpretation. The full extent of the range in a ϕ, ψ plot is $\pm 180°$ for each axis, with 0 as the midpoint of each.

As shown in Figure 3.3, for each transition, the dihedral angle data was calculated with respect to time and with respect to angles PHI vs PSI (ϕ vs ψ). The left side was calculated for the transition from compact to extended form, while the right side was calculated for the transition from extended to compact form. The upper left panel for the transition from compact to extended conformation shows ϕ vs ψ, and indicates the highest density between $[-150, -50]$ ϕ and $[-180, 50]$ ψ. The only other cluster in significant density appears at $[\phi, \psi]$ of $[-30, 150]$. The interpretation of the positions and size/shape of the most significant dense clusters is that of the Serine and Histidine (SER and HSD) residues, which frequently comprise turn regions and are also residues 108-109. However, the bulk of the main cluster indicates a predominately α-helix structure. This may indicate the possibility of residues in the loop/turn region having the capability of assuming helical structure, or vice verse, with the latter being most likely.

The lower left panel for the same transition with respect to time shows very little fluctuation in range until the last 30 ps or so, where the highest deviation is observed. This is with the exception of PSI 109, which backbone neighbors turn residues Glycine and Asparagine. This may indicate that more movement takes place in residues 109-112, particularly between 60-120 ps.

The upper right panel for Figure 3.3 shows the dihedral angle data as calculated for the epsilon subunit transition between the extended and compact form. The

clusters appear as the inverse of the transition from compact to extended, and may even occupy regions bordering disallowed regions in the lower half and upper right quadrant. However, this conflicts with the given locations for clusters indicative of Glycine, which has less absolute defined cluster locations and is the most likely to occupy disallowed regions. [28,29] In that event, it would be critical for the placement of a Glycine in the hinge region to promote the highest range of motion during a transition.

The lower right panel of Figure 3.3 shows the dihedral angle data with respect to time for the transition from the extended to the compact form. It too appears the inverse of the data for the opposite transition. The bulk of the time is spent in avoidance of absolute value acute angles. Here, the significant deviation originates from the PHI 108-109 angles and PSI 108. This may indicate that the reverse transition from the extended to the compact state is more dependent upon the Serine-Histidine (SER-HSD) backbone connectivity closer to the NTD than the CTD, and reverses for the transition in the opposite configuration.

In order to calculate the helical crossing angle, a selection of representative residues must be selected that remain the most constant. In Figure 3.4, three preliminary selections of core selections for the two helices were made, based on pre-trajectory assignments. The best selection correlates with the data most consistent over the trajectory, in this case the middle selection of residues (92-101) and (115-124). This selection was narrowed to (95-99) and (117-121) over subsequent helical crossing angle studies.

The information given by the helical crossing angle between the two α-helices, at least in the transition between compact and extended conformations, is less decisive. Instead, it shows the first 60 ps as having very little change. A steady decrease occurs between 85-141 ps from 146° to 13°. This most likely corresponds to the transition of the two helices to the extended configuration (Figure 3.1 (d-g)). The remaining peak at 130 ps is the final adjustment between the two helices. However, the data for the reverse transition does not appear to be the ideal of a mirror image. Instead,

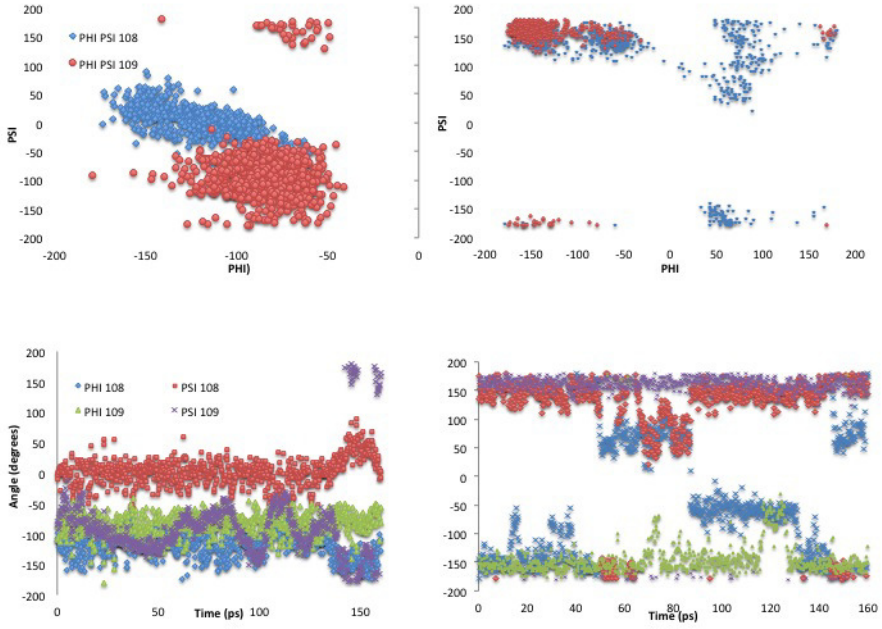


Figure 3.3. Dihedral angle data for transitions to (left) and from (right) ε_χ conformation. Above: PHI vs PSI for residues 108-109, Below: dihedral angles vs time.

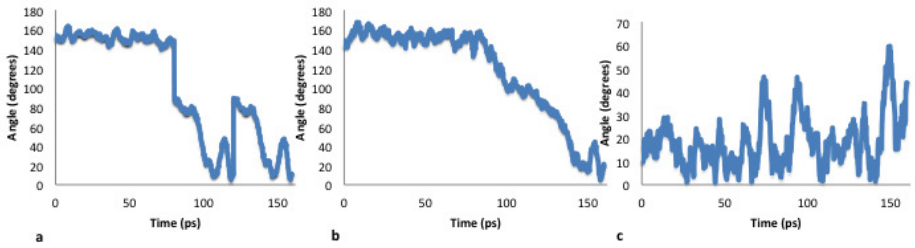


Figure 3.4. Helical crossing angles for the transition from compact to extended form, over three helix atom selections. (Left) Helix 1 residues defined as 90-106, helix 2 residues defined as 111-137. (Center). Helix 1 residues defined as 92-101, helix 2 residues defined as 115-124. (Right) Helix 1 residues defined as 112-126, helix 2 residues defined as 130-137.

it reflects the appearance of residues that undergo unravelling such that residues that once made up outer regions of helices degraded into extensions of surrounding disordered loop regions. A narrower selection of residues for calculating the helical crossing angle would have to be made in order to get a more accurate assessment, but then maintained over the rest of the study. Additionally, if the selection of residues is narrow, the determined angle vectors are more easily skewed.

The angular difference between the two states at $t = 0$ is roughly 120°, or 2/3 the expected value of 180°, most likely due to the difference between the equilibrated starting positions and the non-equilibrated reference positions. However, the angular difference between the two states at the final time 160 ps is 140°, or 3/4 the expected value. Thus, the target positions appear more optimal. In the transition from the extended to the compact state, the most significant change occurs $\tilde{1}$10 ps, corresponding to Figure 3.1 d in the transition from the extended to the compact form. The main interpretation of the angular values is how the compact form should have an angle approximately 180° while the extended form has an angle of approximately 0, as the two helices are roughly above each other instead of next to each other. The values were calculated using the definition provided by Chothia et al. 1981. [30]

In order to better determine the possibility of the Alanine zipper, atomic pairs corresponding to Alanine and Lysine residues on the α helices capable of long-range hydrogen bonding between helices were selected for distance measurement. Figure 3.2 illustrates the relative positions of the selected residues. Cross-pairs were selected to determine if the resulting distances across the entire trajectory crossed over or formed a twist. If present, these events would support the outlined trajectory steps.

An analysis of select atomic distances (see Figure 3.6, left) over the course of the trajectory from the compact to the extended state indicates that the first major transition occurs just after 83 picoseconds between the end of the first disordered region (N-terminal domain, NTD) and the end of the second α-helix (residues 91 OE1 and 135 NZ). A sharp increase in distance occurs between 83 to 94 ps and again between 133-142 ps. However, a sharp decrease occurs between 149-155 ps. In

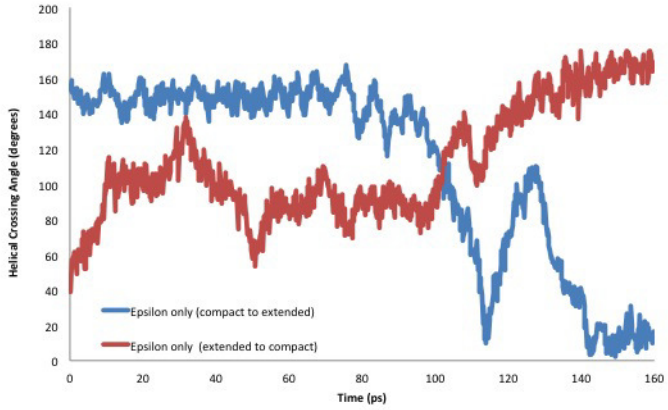


Figure 3.5. Helical crossing angles for isolated epsilon subunit: transitions between compact and extended forms. Helix 1 residues defined as 95-99, helix 2 residues defined as 117-121.

between the two drastic intervals, the increase is moderate. This pattern is similar but attenuated for the distance between two residues centrally located within their respective helices, (97, 124) and (98, 125). The adjacent pairs (97, 125) and (98, 124) paths appear to cross minimally, which indicates that the adjacent pairs would break together instead of sequentially. The line corresponding to the non-CA atoms intersects the clustered pairs $\tilde{8}0$ ps. The last 20 ps shows the convergence of most pairs, while (91, 135) attained the maximum distance from (98, 125), which indicates a distance of $\tilde{2}0$ Å between them–approximately the distance between the start and target structures.

The panel to the right follows the reverse transition from the extended to the compact form. The plot appears as a near reflection of the reverse transition, with the exception of additional fluctuation and decreasing slope from 80-160 ps. The pair (98, 125) is shown to cross (97, 124) between 40-60 ps, which indicates a twisting of the helices. Endpoint positions have converged such that they are consistent with the reference equilibrated target structure.

The table (right) in Figure 3.8 shows majority of the pairs consisting of Alanine, Lysine and Glutamine residues. The table was constructed from the results of tracking all long-range hydrogen bonds between helices and excluding those of neighboring helix turns and adjacent residues, over the entire trajectories. This also determined the exclusion of long-range hydrogen bonds that were not present over both trajectories and therefore not reversible.

Also in 3.8, the table (right) lists the key side chain - main chain long-range hydrogen bonds that are formed or broken between the two helices during conformation transitions. The plot (left) graphically illustrates the pairwise separation with respect to time, which strongly indicates a sequential pairwise breaking or formation of hydrogen bonds. This gives strong support to the Alanine zipper hypothesis as proposed by Uhlin et al. 1997 [2]. While the atomic distance data suggested that long range bonds are broken/formed in clusters or by helix turn, it is the hydrogen

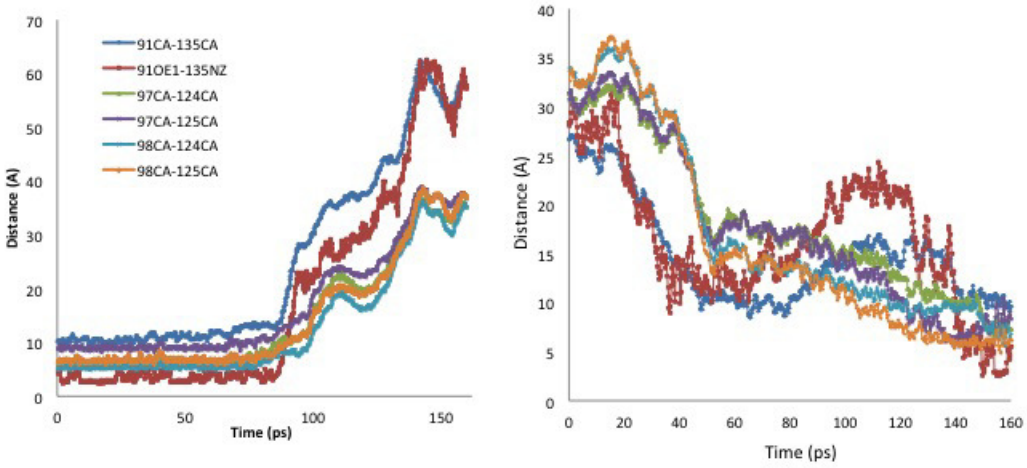


Figure 3.6. From TMD results, select atomic distances over time during transition to (left) and from (right) the extended form. Distances show spacing between different regions of the two helices.

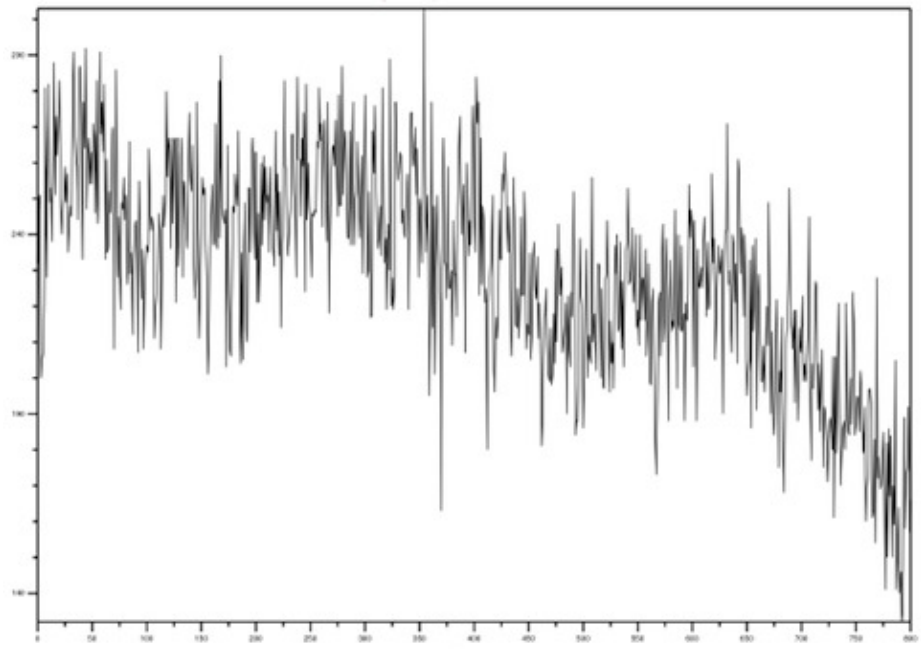
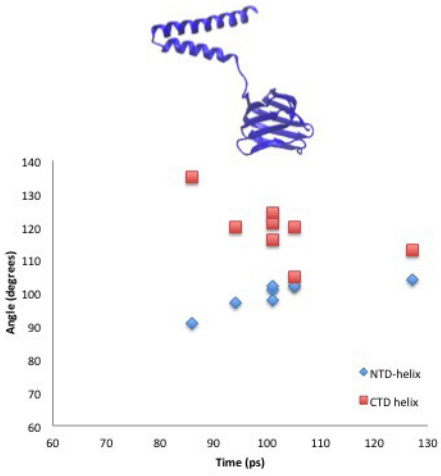


Figure 3.7. VMD generated plot of long range hydrogen bonds by time step of TMD simulation in α helices of transition from compact to extended form.



Time (ps)	Res 1	Resid	Type	Res 2	Resid	Type	Distance (Å)
86	GLU	91	SC O	LYS	135	SC H	2.85
94	ALA	97	SC H	GLY	120	SC O	5.74
101	LYS	98	SC H	LEU	121	SC O	5.88
101	ALA	101	SC H	GLU	116	MC O	5.66
101	GLU	102	SC O	ALA	124	MC H	5.71
105	GLU	102	MC O	GLU	120	MC H	6.95
105	GLU	102	MC O	ILE	105	SC H	4.56
127	HSD	104	SC H	ASP	113	MC O	4.73

Figure 3.8. (Left) Hydrogen bonds linking the two helices that form and break over the duration of the TMD simulation. Residue number vs. time (ps). (Right) Hydrogen bonds linking the two helices that form and break over the duration of the TMD simulation. SC and MC indicate whether the side or main chain participates as donor or acceptor.

bond data that suggests that the formation or breaking of such bonds on a sequential basis allows the separation or bonding to proceed.

In Figure 3.7, the overall number of hydrogen bonds for the entire epsilon subunit decreases over the transition from the compact to the extended form. However, while the trend is reversible for the opposite transition, the trend is overall weak since many of the bonds included in the calculation are still present between close neighboring residues in the helices and between β-strands in the NTD.

3.2.2 Discussion

The simulations featuring only the epsilon segment of proteins 1BSN and 3OAA demonstrate how much of the transition is driven intrinsically by the epsilon subunit. Thus far the model does not account for rotation by the γ subunit, or its corresponding interaction with $(\alpha\beta)_3$ subunits as described by Ma et al. 2002 [11]. Nonetheless, based on the results of the helical crossing angles, atomic distances, dihedral angles and hydrogen bond data over the forward and reverse trajectories, there is evidence to support the Alanine zipper hypothesis. The transition from the extended to the compact form appears less stable than from the compact to the extended conformation, but would be consistent with the scenario that the transition from compact to extended form occurs when there is a need to conserve ATP, while the reverse occurs during a surplus. It would be useful to confirm whether or not the transition of epsilon from the compact to extended form is energetically favorable enough without ATP, while requiring ADP-ATP to transition back from the extended to the compact form. It is also unknown whether hydrogen bonding occurs with surrounding solvent molecules to further stabilize the complex during transitions. The next study of epsilon with the gamma subunit from 3OAA will help determine the affect of gamma on the epsilon segment transitioning between the two conformations.

4. TMD OF ADDITIONAL SUBUNITS

4.1 Gamma, Epsilon Subunits

4.1.1 Experimental

While the Protein Data Bank structure 3OAA contains multiple copies of the $\alpha_3\beta_3\gamma\varepsilon$ assembly, the 1BSN [7, 13] structure contains only ε. In order to examine the behavior of ε in the presence of γ in 1BSN, missing subunits were borrowed from 3OAA and aligned using the program VMD [1] to create a template PDB file. The procedure for obtaining input files, system heating, equilibration, and TMD was performed in the same manner as for ε alone with one exception: in the first 100 K during heating the time step of 2 fs was cut in half, while the nstep was doubled. This ensured a more careful start to the simulation to avoid errors. These settings were restored to previous values for the duration of the simulation. The total computational runtime for the study of $\gamma\varepsilon$ (not counting analysis) required 24.5 hours of parallel processing time across 8 nodes on the Big Red cluster at Indiana University.

4.1.2 Results of the Transitions of Epsilon with Gamma

The results of the helical crossing angle for both transitions was not as cut and dry as the previous epsilon-only results. To improve the results, a more narrow selection of central helix residues would be required. However, if too few residues are selected, again inaccurate results would appear. Figure 4.1 demonstrates a measure of the instability of the endpoint configurations after the transition. The results of the helical crossing angle data would suggest a high amount of fluctuation, particularly

in the time frame between 100-140 ps for both transitions. The convergence of the two values shows a similar value of instability during the most critical time interval for both transition directions. The one redeeming quality is the coincidence of major features at the times corresponding approximately to significant stages in the transition to the extended form of epsilon. Ideally, the start and end point values would have a difference of as close to 180° as possible. The equilibrated starting structure deviates somewhat from the reference starting structure, which may account for some of the difference. At time zero, the angular difference between the two transitions is approximately 140°, or 3/4 of the expected value. After 160 ps, the difference is approximately 120°, or 2/3 of that expected. Again, it is possible that fluctuations at the endpoint (particularly for the transition from the extended to compact state) affected the angle and thus determined the decrease in angular distance between the two transitions. As with the study of the isolated epsilon subunit, the helical crossing angle was calculated according to the method determined by Chothia et al. 1981 [30].

The dihedral angle data was substantially more conclusive. As shown in Figure 4.2, the ϕ, ψ plots are indicative of the amount of flexibility in the hinge disordered loop region. A visual inspection of the trajectories show the transition in both directions as initiated by the hinge region, as opposed to the hook region as observed in the studies of the independent epsilon subunit. Overall, the position of the highest concentration clusters indicates relative stability, an avoidance of "forbidden zones" [28], and behavioral identification with the residues that comprise the disordered loop hinge region, Serine and Histidine (108-109), as indicated by Hovmöller et al. 2002 [29]. However, there are additional regions that indicate the behavior of Glycine and Asparagine, which number as neighboring residues (110-111).

The lower panel data show the time correlation of each calculated angle along the trajectory. The maximum and minimum angle ranges are in agreement with the corresponding panels above. The difficulty remains in the determination of the multiple short excursions of Psi 109 across the upper and lower reaches of the degree range.

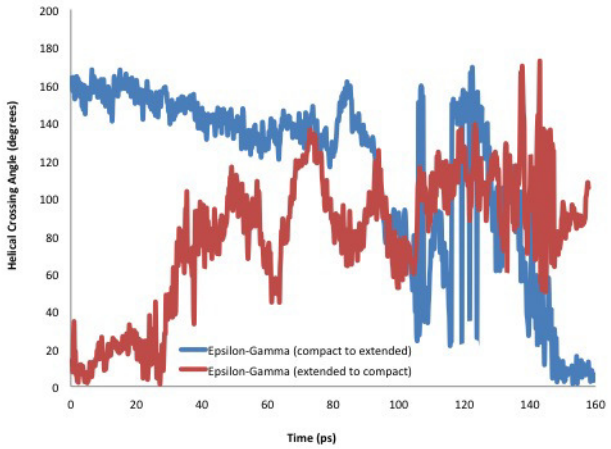


Figure 4.1. Helical crossing angles for epsilon with gamma: transitions between compact and extended forms. Helix 1 residues defined as 95-99, helix 2 residues defined as 117-121.

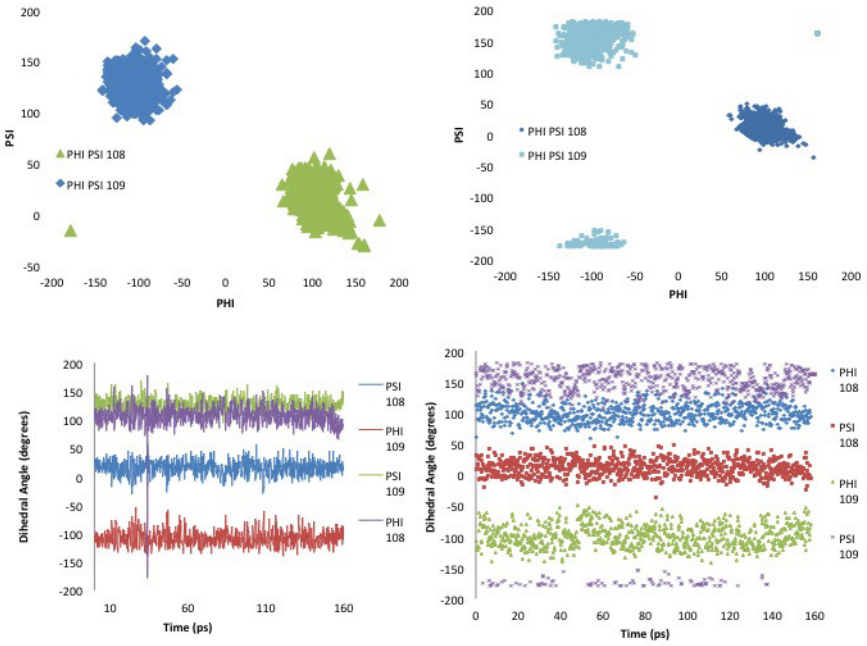


Figure 4.2. Dihedral angle data for transitions to (left) and from (right) ε_χ conformation. Above: PHI vs PSI for residues 108-109, Below: Dihedral angles vs time.

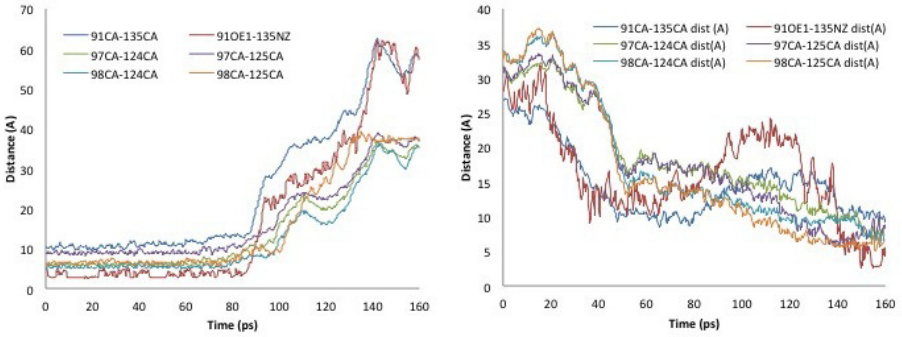


Figure 4.3. From TMD results, select atomic distances over time during transition to (left) and from (right) ε_χ. Distances show spacing between different regions of the two helices.

The study of the atomic distances indicated very similar results to that of the epsilon subunit without gamma. The transition to the extended form still appeared more disordered than that of the reverse transition to the compact form. The amount of non-parallel trajectory occurrences associated with helical cross-over and twisting movements was also comparable to epsilon alone. For instance, the 91CA-135CA pair trajectory crosses all paths between 90-120 ps, while the 91 OE1-135 NZ crosses its CA counterpart after 20, 40, 130 and then 140 ps. The pair of residues 97 CA - 124 CA crosses over that of 97 CA - 125 CA which indicates a twist occurring at approximately 100 ps. Residue pairs 97 CA - 124 CA and 98 CA - 125 CA were mainly convergent, followed by a parallel length preceding a twist at 50 ps.

4.1.3 Discussion

In the epsilon subunit study, it was apparent that the transition from the compact form to the extended form was more reliable than that of the reverse direction. That fact appeared to hold true for the gamma-epsilon complex as well, given the similarity of the atomic distance results between the two studies– in effect, the addition of the gamma subunit did not affect the transitional stability or sequence of events. However

it appeared that improvement in the stability of the epsilon subunit was observed in the dihedral angle study with the addition of the gamma subunit. In addition, the result of the helical crossing angle study after the addition of the gamma subunit appears to also suggest that the same choice of representative α-helices residues does not behave in the same manner as that same selection did in epsilon alone. The helical crossing angle data of the complex suggests that more fraying of helix residues into nearby disordered loop regions occurred for both transitions, and throughout the length of the transition, as opposed to the near completion of the transition. It suggests that in order to have completely stable and reversible transitions, additional stabilizing factors are necessary. It would then be useful to study the presence of all subunits effect on the transitional stability of the epsilon subunit.

4.2 Preliminary Comparison of All Subunits

4.2.1 Experimental

The final investigation for comparison purposes comprised of all chains ($(\alpha\beta)_3\gamma\varepsilon$) in both configurations (ε_χ and ε_C) using RCSB PDB coordinate files 1BSN for the compact conformation and 3OAA for the extended conformation [3, 7, 13] was performed similarly to the previous studies with the exception that the number of atoms prohibited running the parallel installation of CHARMm, and required modification of the input to run in series (the series installation allowed for a larger atomic system, due to architectural differences, even after the upgrade of the operating system to Red Hat Linux 6). Also in order to minimize errors, the time step length was decreased again to 1 fs, while the adjustment to series mode required the number of input steps to increase so that each stage ran for 10 000 - 20 000 time steps and required twice as many stages to avoid the 24 hour runtime limit. Since the heat up and equilibration stages required 8 hours per stage and 10 stages, the preparation for running TMD required $>$ 120 hours for both configurations combined. The TMD phase was again performed at 1 fs per time step for 20 000 time steps per stage, over 8 stages. Each

stage required 18 hours of runtime. The resulting combined computational runtime was $>$ 288 hours on the Quarry cluster at Indiana University.

4.2.2 Discussion

Targeted Molecular Dynamics simulations were performed on the whole complex for both transition directions. A structural investigation of possible interaction between the DELSEED motif of the β subunit and the hook region of the ε subunit was conducted on the basis suggested by Sekiya et al. in 2009 and Futai et al. in 2012 [8,10] of possible interaction between the charged residues of the DELSEED loop and the furthest reach of an extended ε form. While Figure 4.4 indicates the location of the DELSEED loop region, and which residues on both subunits are charged, the preliminary TMD trajectory did not show any interaction with the β DELSEED region. This is consistent with the result from Rodgers and Wilce [4]. In fact, during the expansion of the crown unit, the crown subunits closest to the direction of the ε subunit provided a gap to accommodate ε. However, it is still possible that the DELSEED unit may still interact with epsilon after it has remained in the extended conformation.

4.3 Future Work

At the start of this investigation, the literature on F_1-ATPase provided many possible avenues to explore towards a better understanding of the role of the epsilon subunit in the inhibition of hydrolysis. While several have been investigated during this study, there remain additional questions to explore as a result. The most difficult to determine computationally would be an investigation of how the proton gradient concentration affects the transition between the compact and extended forms of the epsilon subunit. It may be possible to apply techniques used in other software to include changes in temperature and pH, or the addition of ions and ATP.

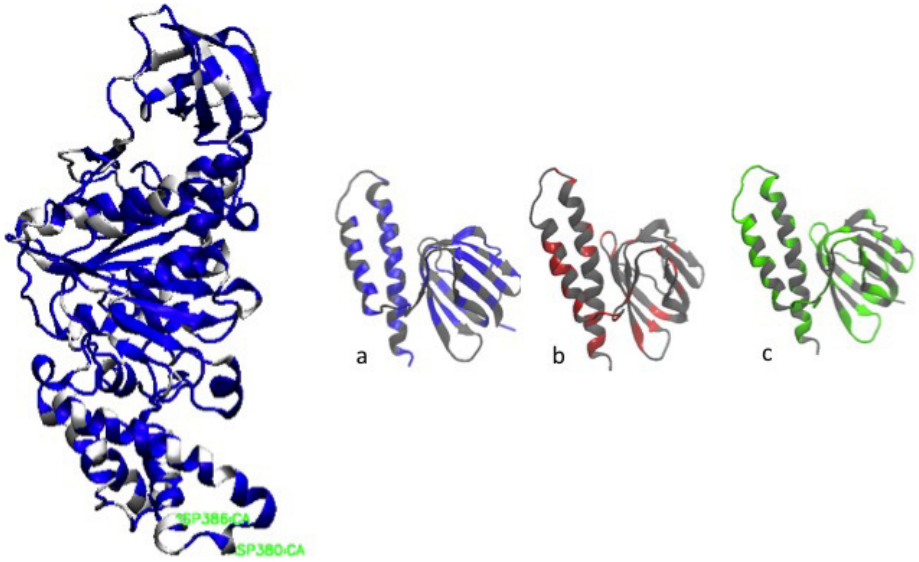


Figure 4.4. (Left) The β subunit of ATPase, in blue with charged residues shown in gray. Labeled in green is the location of the DELSEED motif.
(a) Structure of ε_C, hydrophobic residues shown in blue against structure.
(b) Structure of ε_C, charged residues shown in red against structure.
(c) Structure of ε_C, polar residues shown in green against structure.

Also as described in the literature (e.g., Ma et al. [11]), the gamma subunit is known to rotate as part of the catalytic process. However, the movement observed in this study appeared more of an asymmetrical oscillation. Therefore, it would be useful to measure the angular rotation of the gamma subunit in the presence of epsilon and other subunits. The results would be useful towards the adjustment or confirmation of the current consensus. Additional computational studies may be used to determine the thermodynamics of the transitions, as alluded to by the appearance of unraveling helices during the simulations in this study. To that end, if temperature is a factor in the stability of either form of the epsilon subunit, it would be extremely useful to determine a comparison with the ATP synthase structure from a thermophilic bacterial or Archaea species, since the differences between the two could lead to breakthroughs in medicinal binding target design studies. This is especially critical given the consideration that Cingolani and Duncan 2011 [7] refer to the structural differences between animal mitochondrial ATPase and bacterial ATPase as being significant enough to prevent interference in that of an animal species when targeting an infectious bacteria such as the cause of tuberculosis. Given the rate at which tuberculosis and other bacterial infectious diseases become resistant to current antibiotics, it is worthwhile to devise new strategies in their defeat. With additional modification to resist otherwise destructive temperatures, many more applications are yet possible.

4.4 Conclusions

The significance of the ATP catalytic protein provides a starting point for many interesting and complex questions regarding the structure, function and regulation of ATP catalysis. While a considerable amount of research has been done in this area, there still remain many more critical questions to determine. Comparatively, there has been much less research on the role and behavior of the bacterial epsilon subunit and its equivalent in other ATPase types, and significantly less research that

is computationally based due to the lack of complete coordinate sets available. With the addition of the whole F_1 structure provided by Cingolani and Duncan, it became significantly more feasible to address *in silico* given the number of questions that could now be investigated.

Towards the investigation of a possible transition pathway between the two known configurations, this study conducted a series of energy minimization stages to prepare for a TMD study of the two conformations. This was performed in order of increasing system complexity, from the epsilon subunit alone to the entire complex. The transition was studied from both target directions to investigate the apparent reversibility of the system for every scenario. With the trajectories completed, a series of tests were conducted to measure several characteristics of the transition. Tests included the calculation of dihedral angles for the flexible loop hinge region to determine the intrinsic flexibility and properties due to residues present, the helical crossing angle was measured to determine the progress of the transition between the two conformations, while several key atomic distances were calculated to determine the transition progress and whether twisting occurred. Lastly, data was collected on the time-correlated binding and breaking of long-range hydrogen bonds between different segments of the subunit towards the confirmation or rejection of the Alanine/Lysine zipper mechanism. A brief look in the whole complex based on the trajectory provided a preview of additional questions to be answered, such as the apparent rotation vs asymmetrical oscillation of the gamma subunit, the interaction of the beta subunit DELSEED motif with charged residues of the hook region on the epsilon subunit, and what appears to be the expansion of the diameter of the crown region to accommodate the movement of the epsilon subunit, as indicated by Muench et al. 2011 [9]. All results of the entire complex trajectories have yet to be investigated numerically to the degree that the previous studies included, but that is left to future work and directions.

Overall, several properties intrinsic to the epsilon subunit and some differences in the presence of the gamma subunit were determined. Some intrinsic properties of the

epsilon subunit alone include the higher stability of the compact form over the extended form, and that long-range hydrogen bonds play a significant role in supporting the Alanine zipper hypothesis, and as such it may be proposed that a contributing factor in the initiation of the change in conformation is the concentration of the proton gradient in stabilizing or weakening long-range hydrogen bonds present between the helices and between the adjacent helix and β-sheet NTD. With the addition of the gamma subunit, significant changes resulted. The additional unit influenced the transition dynamics, and the transition appeared slightly more reversible. Additionally, it was observed that the gamma subunit appears to undergo an asymmetrical oscillation as opposed to a steady-state rotation. It is possible that the apparent wobble is stabilized in the presence of the crown subunits. While the trajectory alone of the whole system appeared to demonstrate a crown expansion during both transitions to accommodate the conformational changes in epsilon, additional work is left to the future to determine the transition path of the epsilon subunit as accompanied by all subunits.

LIST OF REFERENCES

LIST OF REFERENCES

[1] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[2] U. Uhlin, G. B. Cox, and J. M. Guss. Crystal Structure of the Epsilon Subunit of the Proton-Translocating ATP Synthase from *Escherichia coli*. *Structure*, 5(9):1219–1230, 1997.

[3] S. Wilkens and R. A. Capaldi. Solution Structure of the Epsilon Subunit of the f_1-ATPase from *Escherichia coli* and Interactions of This Subunit With Beta Subunits in the Complex. *Journal of Biological Chemistry*, 273:26645–26651, 1998.

[4] A. J. W. Rodgers and M. C. J. Wilce. Structure of the Gamma-Epsilon Complex of ATP Synthase. *Nature Structural and Molecular Biology*, 7:1051–1054, 2000.

[5] S. P. Tsunoda, A. J. W. Rodgers, R. Aggeler, M. C. J. Wilce, M. Yoshida, and R. A. Capaldi. Large Conformational Changes of the Epsilon Subunit in the Bacterial f_1f_0 ATP Synthase Provide a Ratchet Action to Regulate This Rotary Motor Enzyme. *Procedings of the National Academy of Sciences of the United States of America*, 98:6560–6564, 2001.

[6] H. Yogi, N. Kajiwara, H. Tanaka, T. Tsukihara, Y. Kato-Yamada, M. Yoshida, and H. Akutsu. Structures of the Thermophilic f_1-ATPase Epsilon Subunit Suggesting ATP-Regulated Arm Motion of Its C-Terminal Domain in f_1. *Procedings of the National Academy of Sciences of the United States of America*, 104:11233–11238, 2001.

[7] G. Cingolani and T. M. Duncan. Structure of the ATP Synthase Catalytic Complex (f_1) From *Escherichia coli* in an Auto-Inhibited Conformation. *Nature Structural and Molecular Biology*, 18:701–707, 2011.

[8] M. Futai, M. Nakanishi-Matsui, H. Okamoto, M. Sekiya, and R. K. Nakamoto. Rotational Catalysis in Proton Pumping ATPases: From *E. coli* F-ATPase to Mammalian V-ATPase. *Biochimica et Biophysica Acta - Bioenergetics*, 1817:1711–1721, 2012.

[9] S. P. Muench, J. Trinick, and M. A. Harrison. Structural Divergence of the Rotary ATPases. *Quarterly Reviews of Biophysics*, 44:311–356, 2011.

[10] M. Sekiya, R. K. Nakamoto, M. K. Al-Shawi, M. Nakanishi-Matsui, and M. Futai. Temperature Dependence of Single Molecule Rotation of the *Escherichia coli* ATP Synthase f_1 Sector Reveals the Importance of the $\gamma-\beta$ Subunit Interactions in the Catalytic Dwell. *Journal of Biological Chemistry*, 284:22401–22410, 2009.

[11] J. Ma, T. C. Flynn, Q. Cui, A. G. W. Leslie, J. E. Walker, and M. Karplus. A Dynamic Analysis of the Rotation Mechanism for Conformational Change in f_1-ATPase. *Structure*, 10:921–931, 2002.

[12] X. Yang, R. Vadrevi, Y. Wu, and C. R. Matthews. Long-Range Side-Chain-Main-Chain Interactions Play Crucial Roles in Stabilizing the $(\beta\alpha)_8$ Barrel Motif of the Alpha Subunit of Tryptophan Synthase. *Protein Science*, 16:1398–1409, 2007.

[13] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28:235–242, 2000.

[14] M. P. Allen. Introduction to Molecular Dynamics Simulations. In N. Attig, K. Bindar, H. Grubmueller, and K. Kremer, editors, *Computational Soft Matter: From Synthetic Polymers to Proteins*, volume 23. NIC Series, Bon. and Germany, 2004.

[15] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A Program for Macromolecular Energy and Minimization and Dynamics Calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.

[16] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.

[17] K. Ramachandran. Molecular mechanics. In K. Ramachandran, G. Deepa, and K. Namboori, editors, *Computational Chemistry and Molecular Modeling*. Springer, 2008.

[18] G. Chen. *Molecular Dynamics Simulations*. Oxford University Press, New York, New York, 2005.

[19] H. Huang, E. Ozkiriuli, and C. B. Post. Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions. *Journal of Chemical Theory and Computation*, 5:1304–1314, 2009.

[20] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1):98–103, July 1967.

[21] R. W. Hockney. The Potential Calculation and Some Applications. *Methods of Computational Physics*, 9:135–211, 1970.

[22] M. A. Cuendet and W. F. van Gunsteren. On the Calculation of Velocity-Dependent Properties in Molecular Dynamics Simulations Using the Leap-Frog Integration Algorithm. *Journal of Chemical Physics*, 127:184102–184110, 2007.

[23] A. van der Vaart and M. Karplus. Simulation of Conformational Transitions by the Restricted Perturbation-Targeted Molecular Dynamics Method. *Journal of Chemical Physics*, 122:114903–1–114903–6, 2005.

[24] U. Haberthuer and A. Caflisch. FACTS: Fast Analytical Continuum Treatment of Solvation. *Journal of Computational Chemistry*, 29:701–715, 2008.

[25] J. Schlitter, M. Engels, P. Kruger, E. Jacoby, and A. Wollmer. Targeted Molecular Dynamics Simulation of Conformational Change–Application to the T-R Transition in Insulin. *Molecular Simulations*, 10:291–308, 1993.

[26] W. Im, M. S. Lee, and C. L. III. Brooks. Generalized Born Model with a Simple Smoothing Function. *Journal of Computational Chemistry*, 24:1691–1702, 2003.

[27] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 1977.

[28] D. Pal and P. Chakrabarti. On Residues in the Disallowed Region of the Ramachandran Map. *Biopolymers*, 63:195–206, 2001.

[29] S. Hovmöller, T. Zhou, and T. Ohlson. Conformations of Amino Acids in Proteins. *Acta Crystallographica, Section D: Biological Crystallography*, 58:768–776, 2002.

[30] C. Chothia, M. Levitt, and D. Richardson. Helix to Helix Packing in Proteins. *Journal of Molecular Biology*, 145:215–250, 1981.